New opportunities in ecological sensing using wireless sensor networks

Scott L. Collins[1], Luís M. A. Bettencourt[2], Aric Hagberg[2], Levi Larkey[2], Renee F. Brown[1], D.I. Moore[1], Greg Bonito[1,4], Kevin A. Delin[3], Shannon P. Jackson[3], David W. Johnson[3], Scott C. Burleigh[3], Richard R. Woodrow[3], and J. Michael McAuley[3]

[1]Department of Biology, University of New Mexico, Albuquerque, NM 87131

[2] T-7, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545

[3]NASA/Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109-8099

[4]Present address: Department of Biology, Duke University, Durham, NC 27708

Abstract

Measuring environmental variables at appropriate temporal and spatial scales remains a significant challenge in ecological research. New developments in wireless sensors and sensor networks will free ecologists from a wired world and revolutionize our ability to study ecological systems at relevant scales. Sensor networks can, however, also analyze and manipulate the data they collect, which moves data processing from the end user to the sensor network itself. Such embedded processing will allow sensor networks to perform data analysis procedures, identify outlier data, alter sampling regimes, and ultimately control experimental infrastructure. We illustrate this capability using a wireless sensor network, the Sensor Web, in a study of microclimate variation under Chihuahuan Desert shrubs. Using Sensor Web data we developed simple analytical protocols for assessing data quality "on-the-fly" that can be programmed into sensor networks. The ecological community can influence the development of environmental sensor networks by working across disciplines to infuse new ideas into sensor network development.

Introduction

Ecologists struggle to measure complex environmental variables that change rapidly in space and time. To date, environmental monitoring and measurement have been limited by methodology, particularly the types of field-based sensors available to ecologists, their costs, and the constraints imposed by the need to physically wire sensors to stationary data loggers. These constraints lead to suboptimal placement of a few sensors within reach of data loggers rather than in locations that optimize measurement of the variable of interest. The development of small, inexpensive wireless sensors (e.g., Johnston et al. 2004) coupled with the widespread availability of low-cost wireless data transmission infrastructure (e.g., Peterson et al. 1995, Atkins et al. 2003) will free us from a wired world and revolutionize our ability to measure environmental variables at appropriate spatial and temporal scales (Martinez et al. 2004, Porter et al. 2005).

Although ecologists are increasingly aware of the power of sensor networks and are involved in the development of new environmental sensors (Palmer et al. 2004), discussion thus far has focused on how such networks will increase our ability to gather data at spatial and temporal scales appropriate for understanding regional ecological phenomena (Porter et al. 2005). Although environmental sensor networks will revolutionize our ability to measure and monitor the environment; they simultaneously pose huge data analysis and management challenges new to most ecologists. Indeed, sensor networks have the capacity to collect and store data at rates that can overwhelm the analytical capability of many users. To date, little attention in the ecological literature has been devoted to explaining the capability of sensor networks to not only collect mind-boggling amounts of data, but also to process, analyze and summarize data in passive ecological monitoring contexts and in active experimental settings.

One approach to this data richness problem is to conceptually reduce a sensor network and associated cyberinfrastructure to three simplified components: (1) the sensor, which is measurement specific, (2) a sensor network that gathers and transmits sensor data, and (3) the end user who analyzes and interprets the data with a particular question in mind. All subcomponents are linked by cyberinfrastructure, including hardwire transmission networks (e.g., the Internet), computers, data archives, and analytical and graphical software. In our stylized decomposition the sensor and the user are problem specific, whereas the sensor network is generalizable across different applications. However, although we traditionally think of data processing as occurring at the user end of the transmission sequence, many sensor networks have the capacity for embedded computing, an important capability that should be exploited by ecologists (Delin and Jackson, 2000, Estrin et al. 2003). Taking advantage of this technology may require shifts in experimental design towards distributed and more real time data screening and analysis, and ultimately adaptive experimentation (Cook et al. 2005), hypothesis formulation, and testing. In short, the network itself produces knowledge from data (Delin et al. 2005).

In this article, we illustrate the potential power of sensor networks to function beyond the acquisition of large, complex data streams. To do so, we briefly present results from an ongoing experiment at the Sevilleta Long-term Ecological Research (LTER) site that uses a wireless sensor network, the Sensor Web, to measure microclimate beneath different species of native desert shrubs. Specifically, we describe the conceptual framework and experimental context in which we are using wireless sensor technology at the Sevilleta. We then highlight the fundamentals of Sensor Web technology, and describe how the sensor network itself can be used for data quality assurance and quality control (QA/QC), data manipulation, and eventually *actuation* – the explicit control of experimental infrastructure based on *in-situ* data processing.

Islands of fertility

Aridland ecosystems worldwide are undergoing dramatic changes in response to a variety of environmental drivers, including rising levels of atmospheric CO_2 (Archer et al. 1995), increased climate variability (Loik et al. 2004), increased atmospheric nitrogen deposition (Fenn et al. 2003), overgrazing (Archer et al. 1988), and changes in natural disturbance regimes (van Auken 2000). One consequence of environmental change in many semiarid regions worldwide is desertification, degradation from the conversion of C_4-dominated grassland to C_3-dominated woodland (Geist and Lambin 2004). Desertification has significant ecological consequences including altered surface and subsurface hydrology (Bhark and Small 2003), reduced biodiversity (Briggs et al. 2005), reduced capacity to retain nutrients (Welter et al. 2006), altered carbon storage capacity (Jackson et al. 2002) and altered soil resource heterogeneity (Schlesinger et al. 1990, 1996) as resources are increasingly concentrated in "islands of fertility" beneath shrub canopies.

Aridland plant communities are characterized by relatively distinct patches of vegetation with intervening bare areas of soil (Schade et al. 2004, Peters et al. 2006, Fig 1a). The original island of fertility model focused on how the local distribution of soil resources changed from relatively uniform to being increasingly concentrated beneath plant canopies. In fact, soil resources are significantly higher beneath grass and shrub canopies compared to bare soil patches at the Sevilleta (Keift et al. 1998). Shrub encroachment not only alters the distribution of soil resources but it may also affect local microclimate. At the Sevilleta we asked the question, “Are all islands of fertility equal?” We were particularly interested in determining how microclimate differed beneath three common native shrub species, a semi-evergreen shrub – creosotebush (*Larrea tridentata*), a small deciduous tree - honey mesquite (*Prosopis glandulosa* var. *torreyana*), and an evergreen shrub - one-seeded juniper (*Juniperus monosperma*). Although this is a relatively straightforward experiment, we chose this design to learn more about how different species modify their local environments, and to assess the longevity and durability of an environmental sensor network in a relatively harsh environment. In the process we gathered extensive data streams that can be used as test beds for embedded data harvesting algorithms and estimation of data error rates within a long-running sensor network.

Sensor Web

Advances in science often begin with the development of new technologies and instrumentation. The integration of 'off-the-shelf' technologies, including microcomputers, sensors, integrated radio chips, and the Internet, allow the development of instrumentation that can scale across space and time, introducing a new paradigm to how we instrument and analyze the environment. The Sensor Web for Ecological Explorations in Terrestrial Systems (SWEETS) project is a collaboration between the Sevilleta LTER and NASA's Jet Propulsion Laboratory (JPL) in which we are exploring the use of an *in situ* sensor network, the Sensor Web, in ecological research to measure canopy microclimate under different species of desert shrubs.

The Sensor Web, developed at JPL (http://sensorwebs.jpl.nasa.gov/), is an amorphous network of spatially distributed sensor platforms (pods) that wirelessly communicate with each other (Fig 1b). This amorphous architecture is unique since it is both synchronous and router-free, making it distinct from Internet-inspired network schemes. The architecture allows every pod to share data with every other pod throughout the network at each measurement cycle. These data sharing protocols in the Sensor Web provide a powerful system for embedded data processing within the sensor network itself.

To measure microclimate variation at the Sevilleta, in October 2003 we placed three Sensor Web pods (v3.1) each in randomly selected open areas, and under the east side canopy of three individuals of each shrub species (12 pods in total) arrayed along a 300 m transect. An additional pod served as a data relay and a 14^{th} pod served as the mother pod which is connected to a laptop that contains the database and serves as a bidirectional portal into the system via the Internet. Sensors on each pod measure soil temperature at 1 and 10 cm depths, soil moisture at 10 cm depth, relative humidity, air temperature and light. After each measurement interval (5 minutes) the pods wirelessly relay their data to all neighboring pods, which allows all pods to construct a synchronous picture of all measurements across the network. SWEETS data are available at http://sev.lternet.edu/research/SWEETS/index.html.

Microclimate differences

A complete analysis of microclimate differences between shrubs species and open areas is not possible or appropriate here. Rather, we provide example data streams (Fig 2a,b) typical of summer and winter conditions, and an analysis of three midsummer microclimate variables (Fig 3) to illustrate the data generated by the Sensor Web (including occasional outlier data), and the potential for using sensor networks in ecological research. Clear differences in soil temperature at two depths and light availability occur between shrub canopies and open areas. In both winter and summer, daily temperature oscillations in bare areas are greater than under creosotebush. Surprisingly, maximum soil temperatures beneath some shrubs are actually higher than in open areas. This results from differences in soil albedo within different microenvironments. Beneath shrubs the soil is covered by organic matter, which darkens the surface and increases heat absorption, particularly during mid-summer. As a result, average daily maximum shallow soil temperature under juniper during July was significantly higher than in open areas or under the canopy of creosotebush and to a lesser extent under mesquite (Fig 3). Nighttime temperature minima were slightly lower under shrubs than in open areas, and maximum daily light levels were lowest under creosotebush. Clearly, not all resource islands are equal which likely has implications for the distribution and abundance of plant and animal species associated with resource islands in aridland environments.

Post hoc analysis of data quality from a large, continuous, ongoing data stream is challenging. Moreover, the deployment of larger and more complex sensor networks will yield huge and ever-growing data sets which will increase the need to automatically screen and analyze large data sets in efficient and timely ways. Such challenges suggest the need for analytical solutions that automatically perform statistical analyses on the fly. These features create the opportunity for validation of ecological hypotheses in real time by the instrument itself, shifting the burden of data analysis and its logistical costs and delays away from the researcher and onto the sensor network. For example, because it is a temporally synchronized, spatially distributed instrument, the Sensor Web has a unique global data sharing protocol among pods that allows for analytical procedures to be easily programmed into each pod so that data quality can be assessed during every measurement interval and data summaries can be generated at any desired measurement interval.

Embedded processing

In the future, large sensor networks will measure multiple environmental variables at short time intervals and operate over vast areas for years. Although the Sensor Web cluster deployed at the Sevilleta has only a modest number of pods, research platforms, such as NEON, envision sensor networks with hundreds or even thousands of sensors. As sensor networks grow to offer better spatial coverage and include remote areas, problems with data quality assessment, storage, retrieval and manipulation will increase quickly outstripping traditional human resources dedicated to offline analysis. Therefore, shifting portions of data analysis from the user to the network itself will be not just a matter of convenience, but a very practical necessity (Delin and Jackson, 2000; Delin 2002, Larkey et al. submitted).

A simple first step in data processing is to identify and eliminate erroneous sensor readings which may occur for many reasons including the occasional sensor measurement or data transmission error. Even if data error rates are very small (e.g.,

<0.001%), large sensor networks with frequent data acquisition protocols will generate potentially hundreds or thousands of error measurements annually. For example, the 12-pod Sensor Web array at the Sevilleta makes six environmental measurements at each node every five minutes yielding more than 12.6 million data points per year. Even if data error rates are exceptionally low, say 1 in 10,000 then about 1,260 data values are potentially erroneous each year. The number of data errors will only increase with the size and complexity of a sensor network. Such small error frequencies could potentially affect overall data quality, possibly creating erroneous action in autonomous systems, as well as reduce the reliability of comparative analyses.

One procedure to eliminate errors is to create an expected probability distribution of the sensor measurement variables. The problem of adopting such a strategy with environmental data is that most data distributions, for example air temperatures or soil water content, change over daily and seasonal cycles. However, a simple and practical way to obtain a more or less stationary short-term estimate amenable to standard statistical inference is to calculate differences among the same measurement variables (e.g., temperature) that are recorded at each time interval. In most sensor arrays, many environmental variables being measured are spatially coherent over relatively large scales, and statistical differences between neighboring sensors are small and their variance bounded. Errors, on the other hand, usually show much larger variation than correct data values. This translates into well behaved probability distributions for the *differences* between measurements of the same variable (Box 1). Such data properties permit the use of an effective data quality assurance scheme that is simple enough to be embedded within a sensor network where all data are shared at each measurement interval. Events that do not fall in the body of the distribution by some predetermined level of statistical confidence can be excluded and identified as errors. Missing values can also be inferred in a similar way (Larkey et al. submitted). In practice such difference distributions can be constructed for each measured quantity and each pair of neighboring sensors, or for averages of multiple sensors in comparable habitats. In this way, sensor networks can be configured to produce knowledge from the raw data, rather than just providing a passive stream of data to the end user.

Data analysis design choices must be balanced between the simplest need for data quality assurance, network memory and processing capabilities, and further treatment of data. For example, the simplest quality assurance algorithm demands no storage but does require several nearby neighbors, as well as complete data sharing among sensor nodes at each measurement interval. It computes the ensemble of differences between all neighbors of a given sensor node and then determines if a given difference is an outlier by comparing it to the average and standard deviation (see Box 1) of difference values. If it is determined to be anomalous the error is assigned to the corresponding measurement and archived with that information in the database.

This strategy for data quality assurance, a vital first step in data processing, yields a constant statistical summary which can be reported and stored in a permanent database along with the raw data. This allows immediate statistical comparisons of data among contrasting environments or time intervals, or between treatment and control areas in ecological experiments. The logic of shifting data analysis to the network allows the sensor network not only to detect naturally anomalous events corresponding to errors, but also to identify important environmental dynamics, such as rainfall pulses in arid environments, which lead to rapid, spatially coherent changes in multiple environmental variables (Potts et al. 2006). Additional algorithms based on quality assurance calculations can then be implemented to increase or decrease sampling frequency in response to environmental triggers. Such adaptive sampling algorithms can be used to reduce data collection rates at times when little change is occurring, and then rapidly increase it whenever necessary to capture changes in environmental conditions at high spatial and temporal resolution. It can also be used to trigger sampling whenever conditions are judged to be sufficiently interesting (Delin 2002).

One of the least discussed aspects of sensor networks in the ecological literature is the potential for *actuation*, the use of autonomous determination of physical changes in the environment to control experimental infrastructure. For example, at the Sevilleta LTER we recently established a multifactor environmental change experiment in which we manipulate nighttime temperatures, winter rainfall, and nitrogen deposition to determine their individual and combined effects on creosotebush encroachment into grassland (Fig. 4). Our nighttime warming treatment is imposed by using lightweight aluminum fabric shelters mounted on rollers similar to a window shade that are drawn across the plots each night to reduce heat loss and elevate nighttime air temperatures. The shelters roll up again each morning. We plan to use our error detection and data summary algorithms described here in this experiment to summarize nightly treatment effects and generate statistical summaries of outlier values that we can use to detect and quickly repair deployment failures. Ultimately, we plan to use embedded processing to calculate statistical differences among treatments and eventually develop *in situ* algorithms to deploy the warming apparatus and to automate the winter rainfall treatments.

Outlook

Wireless sensor networks have tremendous potential in environmental research. Ecologists are becoming increasingly aware of the capability of these networks to collect multiple point measures of ecological variables at high temporal frequencies across vast spatial scales. Development and use of inexpensive, long-lasting sensor networks will increase our ability to conduct research at scales relevant to environmental grand challenges (NRC 2001, 2003). Yet environmental sensor networks offer a far greater potential than simply switching from a wired to a wireless world. Wireless sensor networks with embedded microprocessors can be programmed to assess data quality, modify sampling regimes, and ultimately activate ecological infrastructure. The development and optimal use of such sensor networks will require a multi-disciplinary effort between ecologists, engineers, computer scientists and statisticians to take full advantage of a technology that is likely to revolutionize not only data collection, but also data processing, analysis and manipulation of experimental infrastructure. Because sensor network technology is still maturing, the ecological community is in a unique position to influence the growth of this technology by working across disciplines to infuse new ideas into wireless sensor network development.

Acknowledgements

We thank Jennifer Johnson and Ken Eack for helpful comments and discussions. This research was supported by NSF grant DEB-0217774 for Long-term Ecological Research. The Sensor Web was constructed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

Literature cited

Archer SR, Scifres C, Bassham, CD and Maggio R. 1988. Autogenic succession in a subtropical savanna: Conversion of grassland to thorn woodland. *Ecol Monogr* **58**: 111-27.

Archer SR, Schimel DS and Holland EH. 1995. Mechanisms of shrubland expansion: Land use, climate or CO_2? *Climatic Change* **29**: 91-99.

Atkins DE, Droegemeier KK, Feldman S, Garcia-Molina H, Klein ML, Messerschmitt DG, Messina P, Ostriker JP and Wright MH. 2003. Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panelon Cyberinfrastructure. *www.nsf.gov/cise/sci/reports/atkins.pdf*).

Bhark EW and Small EE. 2003. The relationship between plant canopies and the spatial patterns of infiltration in shrubland and grassland of the Chihuahuan Desert, New Mexico. *Ecosystems* **6**:185-96.

Briggs JM, Knapp AK, Blair JM, Heisler JL, Hoch GA, Lett MS and McCarron JK. 2005. An ecosystem in transition: Causes and consequences of the conversion of mesic grassland to shrubland. *BioScience* **55**: 243-54.

Cook WM, Casagrande DG, Hope D, Groffman PM and Collins SL. 2004. Learning to roll with the punches: Adaptive experimentation in human-dominated systems. *Front Ecol Environ* **2**: 467-74.

Delin KA 2002. The Sensor Web: A macro-instrument for coordinated sensing. *Sensors* **2**: 270–85.

Delin KA and Jackson SP. 2000. Sensor Web for in situ exploration of gaseous biosignatures. *IEEE Aerospace Conf Proc* **7**:465-72.

Delin KA, Jackson SP, Johnson DW, Burleigh SC, Woodrow RR, McAuley JM, Dohm JM, Ip F, Ferre TPA, Rucker DF and Baker VR. 2005. Environmental studies with the Sensor Web: Principles and practice. *Sensors* **5**: 103-17.

Estrin D, Michener W, and Bonito G. 2003. Environmental cyberinfrastructure needs for distributed sensor networks: a report from a National Science Foundation Sponsored Workshop, Scripps Institute of Oceanography (www.lternet.edu/sensor_report).

Fenn ME, Baron JS, Allen EB, Reuth HM, Nydick KR, Geiser L, Bowman WD, Sickman JO, Meixner T, Johnson DW, and Neitlich P. 2003. Ecological effects of nitrogen deposition in the western United States. *BioScience* **53**: 404-20.

Geist HJ and Lambin EF. 2004. Dynamic causal patterns of desertification. *BioScience* **54**: 817-29.

Jackson RB, Banner JL, Jobbagy EG, Pockman WT and Wall DH. 2002. Ecosystem carbon loss with woody plant invasion of grasslands. *Nature* 623-26.

Johnston CA, Groffman P, Breshears DD, Cardon ZG, Currie W, Emanuel W, Guadinski J, Jackson RB, Lajtha K, Nadelhoffer K, Nelson D, Jr, Post WM, Retallack G and Wielopolski L. 2004 . Carbon cycling in soil. *Front Ecol Environ* **2**: 522-28.

Kieft TL, White CS, Loftin SR, Aguilar R, Craig JA and Skaar DA. 1998. Temporal dynamics in soil carbon and nitrogen resources at a grassland-shrubland ecotone. *Ecology* **79**: 671-83.

Larkey L, Bettencourt LMA, Hagberg A. 2006. In situ data quality assurance for environmental applications of wireless sensor networks. *ACM Transactions on Sensor Networks*: submitted.

Loik, ME, Breshears DD, Lauenroth WK, Belnap J. 2004. A multi-scale perspective of water pulses in dryland ecosystems: climatology and ecohydrology of the western USA. *Oecologia* **141**:269-81.

Martinez K, Hart J and Ong R. 2004. Environmental Sensor Networks. *IEEE Computer* **37**: 50-6.

National Research Council. 2001. Grand Challenges in Environmental Sciences. National Academy Press, Washington, DC.

National Research Council. 2003. NEON: Addressing the Nation's Environmental Challenges. National Academy Press, Washington, DC.

Palmer MA, Bernhardt ES, Chornesky EA, Collins SL, Dobson AP, Duke CS, Gold BD, Jacobson RB, Kingsland SE, Kranz RH, Mappin MJ, Martinez ML, Micheli F, Morse JL, Pace ML, Pascual M, Palumbi SS, Reichman OJ, Townsend AR, and Turner MG. 2005. Ecology for the 21st Century: An action Plan. *Front Ecol Environ* **3**: 4-11.

Peters DPC., Gosz JR, Pockman WT, Small EE, Parmenter RR, Collins SL and Muldavin E. 2006. Integrating patch and boundary dynamics to understand and predict biotic transitions at multiple spatial scales. *Landscape Ecol* **21**: 19-33

Peterson RL, Ziemer RE and Borth DE. 1995. Introduction to spread-spectrum communications. Prentice Hall, Englewood Cliffs, NJ.

Porter, J., Arzberger P, Braun H-W, Bryant P, Gage S, Hansen T, Hanson P, Lin C-C, Lin F-P, Kratz T, Michener W, Shapiro S and Williams T. 2005. Wireless sensor networks for ecology. *BioScience* **55**: 561-72.

Potts DL, Huxman TE, Enquist BJ, Weltzin JF and Williams DC. 2006. Resilience and resistance of ecosystem functional response to a precipitation pulse in a semi-arid grassland. *J Ecol* **94**: 23-30.

Schade JD, Sponseller R, Collins SL and Stiles A. 2003. The influence of mesquite canopies on understory vegetation: Effects of landscape position. *J Veg Sci* **14**: 743-50.

Schlesinger WH, Reynolds JF, Cunningham GL, Huenneke LF, Jarrell WM, Virginia RA and Whitford WG. 1990. Biological feedbacks in global desertification. *Science* **247**: 1043-48.

Schlesinger WH, Reikes JA, Hartley AE and Cross AF. 1996. On the spatial pattern of soil nutrients in desert ecosystems. *Ecology* **77**: 364-74.

Van Auken OW. 2000. Shrub invasions of North American semiarid grasslands. A*nn Rev Ecol Syst* **31**:197-215.

List of Figures

Figure 1. (A) Desert grassland vegetation with scattered creosotebush showing the distribution of vegetation, shrub islands and bare soils characteristic of aridland ecosystems. (B) Sensor Web v3.1 pod underneath a juniper tree (see http://sev.lternet.edu/research/SWEETS/index.html for a site map and access to Sensor Web data).

Figure 2. Typical Sensor Web data streams from the Sevilleta LTER site in New Mexico during a five-day period in a) winter, and in b) summer.

Figure 3. Average daily ranges (minimum, maximum) of light flux, air temperature, and shallow soil temperature under creosotebush, mesquite, and juniper, and bare soil for July 2004 at the Sevilleta LTER site in New Mexico.

Figure 4. Photograph of a warming apparatus in a new nighttime warming, winter rainfall, N-deposition experiment at the Sevilleta LTER site in New Mexico.

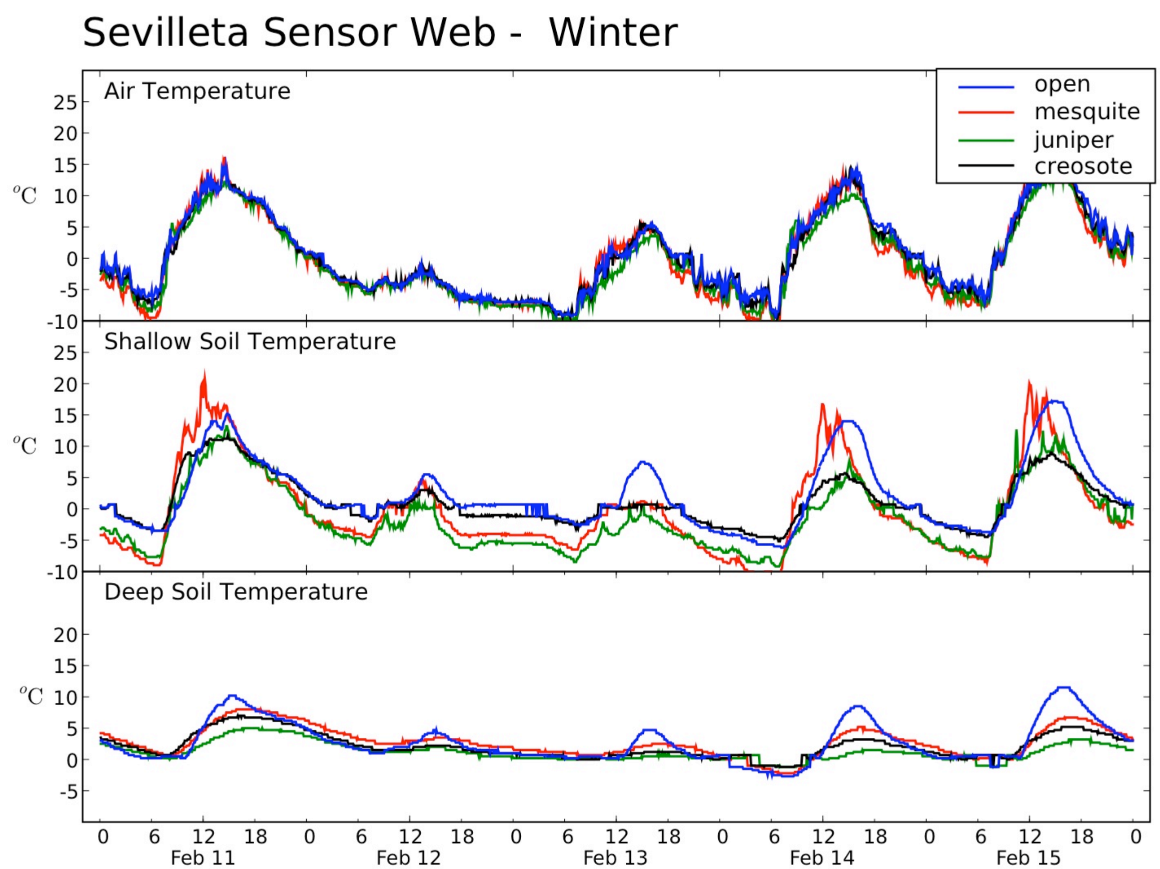
Figure 1

a)

b)

Figure 2

a)


Sevilleta Sensor Web - Winter
Air Temperature
open
mesquite
juniper
creosote
Shallow Soil Temperature
Deep Soil Temperature
^{o}C
25
20
15
10
5
0
-5
-10
0 6 12 18 0 6 12 18 0 6 12 18 0 6 12 18 0 6 12 18 0
Feb 11
Feb 12
Feb 13
Feb 14
Feb 15

b)

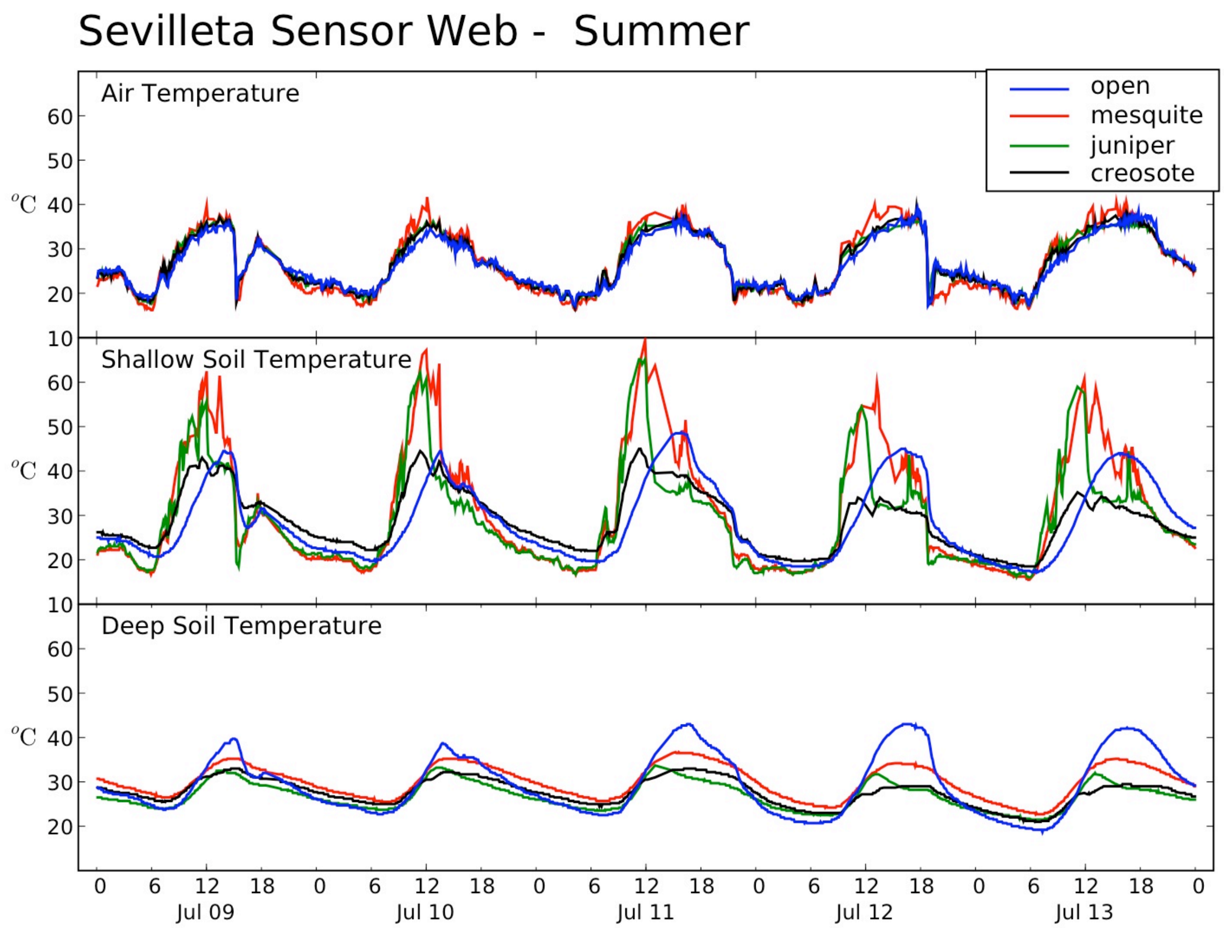
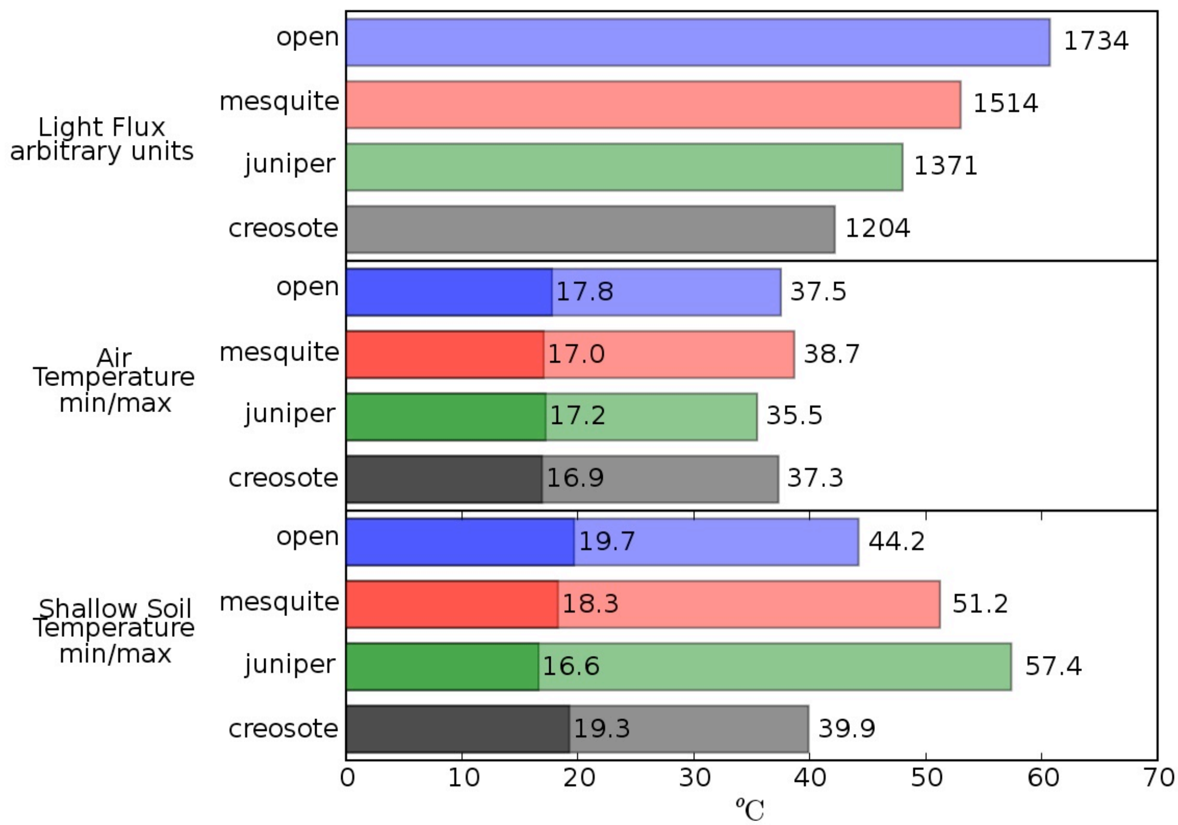

Sevilleta Sensor Web - Summer
Air Temperature
Shallow Soil Temperature
Deep Soil Temperature
open
mesquite
juniper
creosote
^{o}C
10
20
30
40
50
60
0
6
12
18
Jul 09
Jul 10
Jul 11
Jul 12
Jul 13

Figure 3.



Light Flux
arbitrary units
Air
Temperature
min/max
Shallow Soil
Temperature
min/max
open
mesquite
juniper
creosote
1734
1514
1371
1204
17.8
37.5
17.0
38.7
17.2
35.5
16.9
37.3
19.7
44.2
18.3
51.2
16.6
57.4
19.3
39.9
0
10
20
30
40
50
60
70
°C

Figure 4.

Box 1. *In situ* error detection in distributed sensor networks.

In any experiment, errors in sensor measurement or data transmission occur occasionally. In many cases erroneous readings fall within the normal range of daily or seasonal variation and may prove difficult to identify. Box 1, Figure 1 (top right) shows a rare data transmission error from a Sensor Web pod at the Sevilleta LTER site. It is possible to embed algorithms in a pod's processing unit that compare data values among sensors, giving a basis for error detection and for the inference of missing readings. As a consequence, outlier values can be detected and flagged when they occur, and in the process, Sensor Web data are analyzed and summarized on-the-fly.

To do this efficiently, and to accommodate the fact that average values change throughout the day and across seasons, we estimate the probability distribution of differences between a given quantity (e.g. air temperature) measured at adjacent pods $P(\Delta T)$ (see Box Figure 1, bottom right). Measurement errors are identified as point failures that occur with a small probability and typically correspond to large and sudden temperature differences of tens of degrees or more. These can be identified and eliminated at a chosen level of confidence C (say 99%) by standard statistical tests. Using inferred probability distributions from data we determine the probability of observing a difference in measurements between the pod in question and its neighbors that is larger (in absolute value) than the value observed. If this total probability is less than C the measurement is classified as anomalous, otherwise the datum is accepted and stored in the database. As new data values are accepted they can be used to update the probability distribution of valid values on-the-fly. Missing readings at a sensor might also be inferred through knowledge of those of its neighboring pods and the statistical distribution of their differences can also be calculated and noted in the database (Larkey, et al., 2006).

List of Figures, Box 1

Box 1, Figure 1. A data transmission anomaly (top, red circle) in the air temperature measurements at one of the Sensor Web pods can be identified and eliminated via comparison to the measurements transmitted by nearest-neighbor pods (left). The anomalous reading is identified as an outlier (bottom right) in the distribution of differences between the readings at a pod and those of its neighbors at a predetermined level of statistical confidence (Larkey, et al. submitted).

Box 1, Figure 1.